

GenCore version 6.2.1  
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2007, 04:31:56 ; Search time 32216  
Seconds

(without alignments)  
10931.438 Million cell

updates/sec

Title: US-10-535-378-10  
Perfect score: 5095  
Sequence: 1  
tttagagccgggtaggggag.....tgccttctgggcttttcta 5095

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7568541 seqs, 34560148153 residues

Total number of hits satisfying chosen parameters: 15137082

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_env:\*  
2: gb\_pat:\*  
3: gb\_ph:\*  
4: gb\_pl:\*  
5: gb\_pr:\*  
6: gb\_ro:\*  
7: gb\_sts:\*  
8: gb\_sy:\*  
9: gb\_un:\*  
10: gb\_vi:\*  
11: gb\_ov:\*  
12: gb\_htg:\*  
13: gb\_in:\*

AC016814 Mus muscu  
     39    439    8.6    1528    5    AK225446  
 AK225446 Homo sapi  
     c 40    424.4    8.3    258646    12    AC096239  
 AC096239 Rattus no  
     41    424.4    8.3    269847    12    AC113625  
 AC113625 Rattus no  
     42    424.4    8.3    331730    12    AC123384  
 AC123384 Homo sapi  
     43    422.2    8.3    738    5    AY463358  
 AY463358 Homo sapi  
     c 44    411    8.1    438    2    AX381439  
 AX381439 Sequence  
     45    375.8    7.4    463    2    AR266587  
 AR266587 Sequence

# ALIGNMENTS

## RESULT 1

AF110184

LOCUS AF110184 10772 bp DNA linear

PRI 22-JUL-1999

DEFINITION Homo sapiens epithelium-restricted Ets protein ESX  
gene, complete

cds.

ACCESSION AF110184

VERSION AF110184.1 GI:5565858

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;

Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates;

Haplorrhini;

Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 4802 to 9870)

AUTHORS Chang,C.H., Scott,G.K., Kuo,W.L., Xiong,X.,  
Suzdaltseva,Y.,

Park,J.W., Sayre,P., Erny,K., Collins,C., Gray,J.W.  
and Benz,C.C.

TITLE ESX: a structurally unique Ets overexpressed early  
during human

breast tumorigenesis

JOURNAL Oncogene 14 (13), 1617-1622 (1997)

PUBMED 9129154  
 REFERENCE 2 (bases 1 to 10772)  
 AUTHORS Chang,C.H., Scott,G.K., Baldwin,M.A. and Benz,C.C.  
 TITLE Exon 4-encoded acidic domain in the  
 epithelium-restricted Ets  
 factor, ESX, confers potent transactivating capacity  
 and binds to  
 TATA-binding protein (TBP)  
 JOURNAL Oncogene 18 (25), 3682-3695 (1999)  
 PUBMED 10391676  
 REFERENCE 3 (bases 1 to 10772)  
 AUTHORS Chang,C.H., Scott,G.K. and Benz,C.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-NOV-1998) Hematology/Oncology, U.C.S.F.,  
 505  
 Parnassus Ave., San Francisco, CA 94143-1270, USA  
 FEATURES Location/Qualifiers  
 source 1. .10772  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="1q32"  
 misc\_feature 34. .622  
 /note="similar to THC 213038"  
 repeat\_region 921. .1524  
 /rpt\_family="Alu"  
 /rpt\_type=dispersed  
 repeat\_region 2978. .3293  
 /rpt\_family="Alu"  
 /rpt\_type=dispersed  
 CAAT\_signal 4697. .4702  
 /inference="non-experimental evidence, no  
 additional details recorded"  
 TATA\_signal 4735. .4736  
 /inference="non-experimental evidence, no  
 additional details recorded"  
 mRNA join(4777. .4888,5311. .5481,6139.  
 .6360,6526. .6618,  
 6822. .6941,7129. .7218,7364. .7480,8011.  
 .8206,9076. .9872)  
 /product="epithelium-restricted Ets protein  
 ESX"  
 5'UTR join(4777. .4888,5311. .5318)  
 4777 4888

```

exon          4777. .4888
               /number=1
misc_feature  4785. .4901
               /note="putative CpG island"
exon          5311. .5481
               /number=2
CDS           join(5319. .5481,6139. .6360,6526.
.6618,6822. .6941,
               7129. .7218,7364. .7480,8011. .8206,9076.
.9190)
               /note="epithelial-restricted with serine box;
Homo
               sapiens ESX cDNA ORF presented in GenBank
Accession
               Number U66894"
               /codon_start=1
               /product="epithelium-restricted Ets protein
ESX"
               /protein_id="AAD45237.1"
               /db_xref="GI:5565859"

```

```

/translation="MAATCEISNIFSNYFSAMYSSDSTLASVPPAATFGADDLVLT
SNPQMSLEGTEKASWLGEQPQFWSKTQVLDWISYQVEKNKYDASAI DFSRCMDGATL
CNCALEELRLVFGPLGDQLHAQLRDLTSSSSDEL SWIIELLEKDGMAFQEALDPGPF
D
QGSPFAQELLDDGQQASPYHPGSCGAGAPSPGSSDVSTAGTGASRSSHSSDSGGSDVD
LDPTDGKLFPSDGFRDCKKGD PKH GK RKRGRPRKLSKEYWDCLEGGKSKHAPRGTHLW
EFIRDILIHPELNEGLMKWENRHEGVFKFLRSEAVAQLWGQKKKNSNM TYEKL SRAMR
YYYKREILERV DGRRLVYKFGKNSSGWKEEEVLQSRN"

```

```

repeat_region 5773. .6059
               /rpt_family="Alu"
               /rpt_type=dispersed
exon          6139. .6360
               /number=3
exon          6526. .6618
               /number=4
exon          6822. .6941
               /number=5
exon          7129. .7218
               /number=6
exon          7364. .7480
               /number=7

```

```

misc_feature      7401. .7525
                  /note="putative CpG island"
exon              8011. .8206
                  /number=8
repeat_region     8655. .8775
                  /rpt_family="Alu"
                  /rpt_type=dispersed
exon              9076. .9872
                  /number=9
3'UTR             9191. .9872
polyA_signal      9845. .9850
                  /inference="non-experimental evidence, no
additional
                  details recorded"
misc_feature      complement(9952. .10387)
                  /note="similar to THC 209689"
misc_feature      10358. .10772
                  /note="similar to THC 203540"

```

# ORIGIN

```

Query Match          100.0%;  Score 5095;  DB 5;  Length
10772;
Best Local Similarity 100.0%;  Pred. No. 0;
Matches 5095;  Conservative 0;  Mismatches 0;  Indels
0;  Gaps 0;

```

```

Qy          1
TTTAGAGCCGGGTAGGGGAGCGCAGCGGCCAGATACCTCAGCGCTACCTGGCGGAACTGG 60

```

```

|||||
Db          4777
TTTAGAGCCGGGTAGGGGAGCGCAGCGGCCAGATACCTCAGCGCTACCTGGCGGAACTGG 4836

```

```

Qy          61
ATTTCTCTCCCGCCTGCCGGCCTGCCTGCCACAGCCGGACTCCGCCACTCCGGTAGGATT 120

```

```

|||||
Db          4837
ATTTCTCTCCCGCCTGCCGGCCTGCCTGCCACAGCCGGACTCCGCCACTCCGGTAGGATT 4896

```

```

Qy          121
CCCCGCCTGTCATTCCCTAGCCCAGCTCTTGGGAAACTGCAGAGGGGTCCAGAGGATTTG 180

```

```

|||||
Db          4897
CCCCGCCTGTCATTCCCTAGCCCAGCTCTTGGGAAACTGCAGAGGGGTCCAGAGGATTTG 4956

```

Qy 181  
CAGTTCTGAACCTGCACACTCCAGTCTAGGATCTCCGAGCAAGAGCGTAGGTGTCCTGAG 240

|||||

Db 4957  
CAGTTCTGAACCTGCACACTCCAGTCTAGGATCTCCGAGCAAGAGCGTAGGTGTCCTGAG 5016

Qy 241  
GGTCAAAGAACAGAGAGAGATTGTCTCTGGGAAGGCAGAATGGCCATGACGCCGCTAGTC 300

|||||

Db 5017  
GGTCAAAGAACAGAGAGAGATTGTCTCTGGGAAGGCAGAATGGCCATGACGCCGCTAGTC 5076

Qy 301  
TGGCTCCAGGGCCCCAGAGATCTGAGGAGGGAAGCCCAGCTGGAGGCTCCTGTGGTCCTG 360

|||||

Db 5077  
TGGCTCCAGGGCCCCAGAGATCTGAGGAGGGAAGCCCAGCTGGAGGCTCCTGTGGTCCTG 5136

Qy 361  
CCCTGGTCTGAGATCTTGGAGCCCTTCTTGAAGAGACGGTGTCCGCAGAGTTGCTGATCT 420

|||||

Db 5137  
CCCTGGTCTGAGATCTTGGAGCCCTTCTTGAAGAGACGGTGTCCGCAGAGTTGCTGATCT 5196

Qy 421  
TCCTGCCCCTGGGGGCTACTCTTGCCCAGGGTTGGGCAAAGCAGAGTAGCTGGGAGTGTA 480

|||||

Db 5197  
TCCTGCCCCTGGGGGCTACTCTTGCCCAGGGTTGGGCAAAGCAGAGTAGCTGGGAGTGTA 5256

Qy 481  
AGGAGAGGACCCTCGTCCCCTCACCAACCTCATCCTCTCTCCCCCTACCCACAGGTAGCC 540

|||||

Db 5257  
AGGAGAGGACCCTCGTCCCCTCACCAACCTCATCCTCTCTCCCCCTACCCACAGGTAGCC 5316

Qy 541  
TCATGGCTGCAACCTGTGAGATTAGCAACATTTTGTAGCAACTACTTCAGTGCGATGTACA 600

.....

|||||  
Db 5317  
TCATGGCTGCAACCTGTGAGATTAGCAACATTTTGTAGCAACTACTTCAGTGCGATGTACA 5376

Qy 601  
GCTCGGAGGACTCCACCCTGGCCTCTGTTCCCCCTGCTGCCACCTTTGGGGCCGATGACT 660

|||||  
Db 5377  
GCTCGGAGGACTCCACCCTGGCCTCTGTTCCCCCTGCTGCCACCTTTGGGGCCGATGACT 5436

Qy 661  
TGGTACTGACCCTGAGCAACCCCCAGATGTCATTGGAGGGTACAGGTGGGTCTCAGCGGG 720

|||||  
Db 5437  
TGGTACTGACCCTGAGCAACCCCCAGATGTCATTGGAGGGTACAGGTGGGTCTCAGCGGG 5496

Qy 721  
GTGGGATGGGGCACGGAGTGGGAGACAGATCCATCTAAGGGCCTGTTAGACAAATGGGGG 780

|||||  
Db 5497  
GTGGGATGGGGCACGGAGTGGGAGACAGATCCATCTAAGGGCCTGTTAGACAAATGGGGG 5556

Qy 781  
AATAGGCAGGGAGGAGGGTCTCTAGGCAAATTCCAGGGCTAGAGGCTGAGACTTAGTGAC 840

|||||  
Db 5557  
AATAGGCAGGGAGGAGGGTCTCTAGGCAAATTCCAGGGCTAGAGGCTGAGACTTAGTGAC 5616

Qy 841  
TGAGGTGCTGGGGGTTGTGGGGCTGTGACAGGCAGAGGGAGGTGTCAGATACCAGGACAA 900

|||||  
Db 5617  
TGAGGTGCTGGGGGTTGTGGGGCTGTGACAGGCAGAGGGAGGTGTCAGATACCAGGACAA 5676

Qy 901  
GGGTGTTGTGAATGCTACCTCCTGCCCCTACTCTTGGGATGGCTCCAAGGGCTGAGGTGT 960

|||||  
Db 5677  
GGGTGTTGTGAATGCTACCTCCTGCCCCTACTCTTGGGATGGCTCCAAGGGCTGAGGTGT 5736

Qy 961  
GAATCCCCAGTGTGCTCCAGGAATGGGGCTGTGTGGGCTGGGAGTGGTGGCTCACGCCTG 1020

|||||  
Db 5737  
GAATCCCCAGTGTGCTCCAGGAATGGGGCTGTGTGGGCTGGGAGTGGTGGCTCACGCCTG 5796

Qy 1021  
TAATCCCAGCACTTTGGGAGGCTGAGCTGAGCGGATCACCTGAGGTCAAGAGTTCGAGAC 1080

|||||  
Db 5797  
TAATCCCAGCACTTTGGGAGGCTGAGCTGAGCGGATCACCTGAGGTCAAGAGTTCGAGAC 5856

Qy 1081  
CAGCCTAGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAAAAATTTATCCC 1140

|||||  
Db 5857  
CAGCCTAGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAAAAATTTATCCC 5916

Qy 1141  
AGCGTGGTGGTGGGCACCTATAATCCCAGCTACTGGGGAGGCTGACGCAGGAGTATCGCT 1200

|||||  
Db 5917  
AGCGTGGTGGTGGGCACCTATAATCCCAGCTACTGGGGAGGCTGACGCAGGAGTATCGCT 5976

Qy 1201  
TGAACCTGGGAGGTGGAGGTTGCTGTGAGCCGAGATTGTGCCATTGCACCCCAGCCTAGG 1260

|||||  
Db 5977  
TGAACCTGGGAGGTGGAGGTTGCTGTGAGCCGAGATTGTGCCATTGCACCCCAGCCTAGG 6036

Qy 1261  
TGACAGGAGTGAGACTCCATCTCAAAAAAAAAAAAAAAAAAATGGGGCTGTAAGGTCTGCTG 1320

|||||  
Db 6037  
TGACAGGAGTGAGACTCCATCTCAAAAAAAAAAAAAAAAAAATGGGGCTGTAAGGTCTGCTG 6096

Qy 1321  
GGTGGCCTGAGCTGAGCCTGTTTCCCTGCCTGGCCCTTGACAGAGAAGGCCAGCTGGTTGG 1380

|||||



Db 6097  
GGTGGCCTGAGCTGAGCCTGTTTCCCTGCCTGGCCCTTGCAGAGAAGGCCAGCTGGTTGG 6156

Qy 1381  
GGGAACAGCCCCAGTTCTGGTCGAAGACGCAGGTTCTGGACTGGATCAGCTACCAAGTGG 1440

Db 6157  
 GGGAACAGCCCCAGTTCTGGTCGAAGACGCAGGTTCTGGACTGGATCAGCTACCAAGTGG 6216

Qy 1441  
AGAAGAACAAGTACGACGCAAGCGCCATTGACTTCTCACGATGTGACATGGATGGCGCCA 1500

|||||  
 Db 6217  
 AGAAGAACAAGTACGACGCAAGCGCCATTGACTTCTCACGATGTGACATGGATGGCGCCA 6276

Qy 1501  
CCCTCTGCAATTGTGCCCTTGAGGAGCTGCGTCTGGTCTTTGGGCCTCTGGGGGACCAAC 1560

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db               6277  
**CCCTCTGCAATTGTGCCCTTGAGGAGCTGCGTCTGGTCTTTGGGCCTCTGGGGACCAAC** 6336

Qy 1561  
TCCATGCCCAGCTGCGAGACCTCAGTGAGTCCAGGCCCTGGAGGCTGGGGAGCAGCTCC 1620

|||||  
Db 6337  
TCCATGCCCAGCTGCGAGACCTCAGTGAGTCCAGGCCCTGGAGGCTGGGGAGCAGCTCC 6396

Qy 1621  
ACATGTTGAGCTGAGTCGAGTTCAGTGTGGCCGTAGGCAGGCCCTGGAGCTCTGGGCCAG 1680

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db           6397  
ACATGTTGAGCTGAGTCGAGTTCAGTGTTGGCCGTAGGCAGGCCCTGGAGCTCTGGGCCAG 6456

---

Qy 1681  
CTGCACAGCCAGAGAGAGCCCTTGAGGGAGGGATTAGGGGAGTGTGACCCCTTCCTTCCTT 1740

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Dbb                  6457

CTGCACAGCCAGAGAGAGCCCCTTGAGGGAGGGATTAGGGGAGTGTGACCCTTCCTTCCTT 6516

[illegible]

CCTTGTCTCAGCTTCCAGCTCTTCTGATGAGCTCAGTTGGATCATTGAGCTGCTGGAGAAGG 1800

|||||

Db 6517

CCTTGTCTCAGCTTCCAGCTCTTCTGATGAGCTCAGTTGGATCATTGAGCTGCTGGAGAAGG 6576

Qy 1801

ATGGCATGGCCTTCCAGGAGGCCCTAGACCCAGGGCCCTTTGGTGAGAACCCGTTTCTC 1860

|||||

Db 6577

ATGGCATGGCCTTCCAGGAGGCCCTAGACCCAGGGCCCTTTGGTGAGAACCCGTTTCTC 6636

Qy 1861

CTTCCTTCCCCAGCCTGTCTTGTTCCCATCCCTGCCCTCCACAGAGTGCTAGAGATGACC 1920

|||||

Db 6637

CTTCCTTCCCCAGCCTGTCTTGTTCCCATCCCTGCCCTCCACAGAGTGCTAGAGATGACC 6696

Qy 1921

CCCTCCCCAGACTTCTTCCTCCCTCAATTAGAAAAATTGCAGCAGGTCATCAGACCCATG 1980

|||||

Db 6697

CCCTCCCCAGACTTCTTCCTCCCTCAATTAGAAAAATTGCAGCAGGTCATCAGACCCATG 6756

Qy 1981

GGCAGCATCACCTGTCCTGGTCTGGTCCCCTGAGCCCTCTCTGAGTTCTCACCTCCTCTT 2040

|||||

Db 6757

GGCAGCATCACCTGTCCTGGTCTGGTCCCCTGAGCCCTCTCTGAGTTCTCACCTCCTCTT 6816

Qy 2041

CCCAGACCAGGGCAGCCCCTTTGCCCAGGAGCTGCTGGACGACGGTCAGCAAGCCAGCCC 2100

|||||

Db 6817

CCCAGACCAGGGCAGCCCCTTTGCCCAGGAGCTGCTGGACGACGGTCAGCAAGCCAGCCC 6876

Qy 2101

CTACCACCCCGGCAGCTGTGGCGCAGGAGCCCCCTCCCCTGGCAGCTCTGACGTCTCCAC 2160

|||||

Db 6877

C -CC-CCCC-CC-CC -CC-CC-CC-CCCC CCCC -CC-CC C -CC- C CC-C -CC-

CTACCACCCCGGCAGCTGTGGCGCAGGAGCCCCCTCCCCTGGCAGCTCTGACGTCTCCAC 6936

Qy 2161  
CGCAGGTGAGAGCTCTCTCTGGGCCACAACCTCCCTTCCCCGAAGTGTCCCTTGTTCCT 2220

|||||  
Db 6937  
CGCAGGTGAGAGCTCTCTCTGGGCCACAACCTCCCTTCCCCGAAGTGTCCCTTGTTCCT 6996

Qy 2221  
CTGGCTCCCAGCACCATAACTCAGGCCTTCTGGCAGGAACAGGAACAGGCTGGGAAGTGT 2280

|||||  
Db 6997  
CTGGCTCCCAGCACCATAACTCAGGCCTTCTGGCAGGAACAGGAACAGGCTGGGAAGTGT 7056

Qy 2281  
GTCCTGAGAGCCAGCAGCGTGGTTGAACAGAAGGTGGGCCGGCAGGGGACTTACTCTGAC 2340

|||||  
Db 7057  
GTCCTGAGAGCCAGCAGCGTGGTTGAACAGAAGGTGGGCCGGCAGGGGACTTACTCTGAC 7116

Qy 2341  
CCCGCCCCCAGGGACTGGTGCTTCTCGGAGCTCCCACTCCTCAGACTCCGGTGGAAGTG 2400

|||||  
Db 7117  
CCCGCCCCCAGGGACTGGTGCTTCTCGGAGCTCCCACTCCTCAGACTCCGGTGGAAGTG 7176

Qy 2401  
ACGTGGACCTGGATCCCACTGATGGCAAGCTCTTCCCCAGCGGTGAGTCGAGGGAGGTCC 2460

|||||  
Db 7177  
ACGTGGACCTGGATCCCACTGATGGCAAGCTCTTCCCCAGCGGTGAGTCGAGGGAGGTCC 7236

Qy 2461  
CCAAGAGGGCGTCCCATTTAGCAATGCACAGGGGGCCCGGCTCTTCCTGCAGCCTTTTCC 2520

|||||  
Db 7237  
CCAAGAGGGCGTCCCATTTAGCAATGCACAGGGGGCCCGGCTCTTCCTGCAGCCTTTTCC 7296

Qy 2521  
TGTAAGAGGGGCTACTCTCCCTAACTCCCCTCTTGCCCCCTCCTTGACCTTCCACCACCGTC 2580

|||||  
Db 7297  
TG TAGAGGGGCTACTCTCCCTAACTCCCCTCTTGCCCCCTCCTTGACCTTCCACCACCGTC 7356

Qy 2581  
CCCACAGATGGTTTTTCGTGACTGCAAGAAGGGGGATCCCAAGCACGGGAAGCGGAAACGA 2640

|||||  
Db 7357  
CCCACAGATGGTTTTTCGTGACTGCAAGAAGGGGGATCCCAAGCACGGGAAGCGGAAACGA 7416

Qy 2641  
GGCCGGCCCCGAAAGCTGAGCAAAGAGTACTGGGACTGTCTCGAGGGCAAGAAGAGCAAG 2700

|||||  
Db 7417  
GGCCGGCCCCGAAAGCTGAGCAAAGAGTACTGGGACTGTCTCGAGGGCAAGAAGAGCAAG 7476

Qy 2701  
CACGGTGAGCTCCGGGGGCACGTGGGTCCCTCCCTGCGCCGGGCTGAGCGGCTTCCTGGGG 2760

|||||  
Db 7477  
CACGGTGAGCTCCGGGGGCACGTGGGTCCCTCCCTGCGCCGGGCTGAGCGGCTTCCTGGGG 7536

Qy 2761  
CACTGCGGGTTGTTGCAGGTATCCCTTCTCCCGTTTTCTCTGGCCTCCGCATGGCCTTTG 2820

|||||  
Db 7537  
CACTGCGGGTTGTTGCAGGTATCCCTTCTCCCGTTTTCTCTGGCCTCCGCATGGCCTTTG 7596

Qy 2821  
GTAAGGCTGTGCACAAGCTGGGGGCTCTATGGTATCGGTCACCACCTAATTGCAGAGCCT 2880

|||||  
Db 7597  
GTAAGGCTGTGCACAAGCTGGGGGCTCTATGGTATCGGTCACCACCTAATTGCAGAGCCT 7656

Qy 2881  
GGCTTGGTGGTCCTGGAGAGGAGGAGGAAATAAGGCTCCCAGTGGGAGGCTCATGGTACC 2940

|||||  
Db 7657  
GGCTTGGTGGTCCTGGAGAGGAGGAGGAAATAAGGCTCCCAGTGGGAGGCTCATGGTACC 7716

Qy 2941  
AGAGTCCTGTCCACTGACTCCAGTGTCTGTCCACTGACTCCAGTTCTCTCTGCACTTGG 3000

|||||  
Db 7717  
AGAGTCCTGTCCACTGACTCCAGTGTCTGTCCACTGACTCCAGTTCTCTCTGCACTTGG 7776

Qy 3001  
CCACTGTCCTGCCCTCTGGGACACCCTCAATGTGAGGAGGCAGCTGGTGGGTCTTAGGTG 3060

|||||  
Db 7777  
CCACTGTCCTGCCCTCTGGGACACCCTCAATGTGAGGAGGCAGCTGGTGGGTCTTAGGTG 7836

Qy 3061  
GGCTGAGGAGAAAAGCAGTCACTGCAGTACCCGCACAGAGGGCACTGCGGGGTCTCTGGA 3120

|||||  
Db 7837  
GGCTGAGGAGAAAAGCAGTCACTGCAGTACCCGCACAGAGGGCACTGCGGGGTCTCTGGA 7896

Qy 3121  
GAGGCTTGCTGCATGCTGTGGCCAAGTCAGCAGTGCAGTGGGGCGGGCAGGGCTGGCTGG 3180

|||||  
Db 7897  
GAGGCTTGCTGCATGCTGTGGCCAAGTCAGCAGTGCAGTGGGGCGGGCAGGGCTGGCTGG 7956

Qy 3181  
CCTTGGGTGAGAGGGGACACCTGGATGGCAAACCTGATGGAGGCTGGCCTTGACAGCGCCCA 3240

|||||  
Db 7957  
CCTTGGGTGAGAGGGGACACCTGGATGGCAAACCTGATGGAGGCTGGCCTTGACAGCGCCCA 8016

Qy 3241  
GAGGCACCCACCTGTGGGAGTTCATCCGGGACATCCTCATCCACCCGGAGCTCAACGAGG 3300

|||||  
Db 8017  
GAGGCACCCACCTGTGGGAGTTCATCCGGGACATCCTCATCCACCCGGAGCTCAACGAGG 8076

Qy 3301  
GCCTCATGAAGTGGGAGAATCGGCATGAAGGCGTCTTCAAGTTCCTGCGCTCCGAGGCTG 3360

.....

|||||  
Db 8077  
GCCTCATGAAGTGGGAGAATCGGCATGAAGGCGTCTTCAAGTTCCTGCGCTCCGAGGCTG 8136

Qy 3361  
TGGCCCAACTATGGGGCCAAAAGAAAAAGAACAGCAACATGACCTACGAGAAGCTGAGCC 3420

|||||  
Db 8137  
TGGCCCAACTATGGGGCCAAAAGAAAAAGAACAGCAACATGACCTACGAGAAGCTGAGCC 8196

Qy 3421  
GGGCCATGAGGTGAGCTGGCGGCCAGGACCCTCACGATACAGCCGGACATGGGGACAGGC 3480

|||||  
Db 8197  
GGGCCATGAGGTGAGCTGGCGGCCAGGACCCTCACGATACAGCCGGACATGGGGACAGGC 8256

Qy 3481  
GCTCACACTCCCACCGCCCTCTTTCTGGCTGCCACTTGGTTTCTTGCAACAGGGCTGAGT 3540

|||||  
Db 8257  
GCTCACACTCCCACCGCCCTCTTTCTGGCTGCCACTTGGTTTCTTGCAACAGGGCTGAGT 8316

Qy 3541  
CCTTAGAGTGAGGACAACATCTGGGTTGGTCTACTTCATGGATTAAATGACAACATGGAG 3600

|||||  
Db 8317  
CCTTAGAGTGAGGACAACATCTGGGTTGGTCTACTTCATGGATTAAATGACAACATGGAG 8376

Qy 3601  
AAAGTATTAGCCTGGCAGACAGCAGACACAGTGCACTTGAGCTAGCAGCAACATTTCTTG 3660

|||||  
Db 8377  
AAAGTATTAGCCTGGCAGACAGCAGACACAGTGCACTTGAGCTAGCAGCAACATTTCTTG 8436

Qy 3661  
TATCGCCTGTGAGGCTTGTCTCAGGAAGGCACCTGGAGAGTGGGAAAGGGGGCAGGAGC 3720

|||||  
Db 8437  
TATCGCCTGTGAGGCTTGTCTCAGGAAGGCACCTGGAGAGTGGGAAAGGGGGCAGGAGC 8496

Qy 3721  
CGTGCCACCCAGGGCCTGGCTTTCTCCTCGTTGAAGCACTTAGGTTGTTTTCTCTGGG 3780

|||||

Db 8497  
CGTGCCACCCAGGGCCTGGCTTTCTCCTCGTTGAAGCACTTAGGTTGTTTTCTCTGGG 8556

Qy 3781  
CCTCAGTTTCCTCCTGTGTCCAGGAGTACACTAGATCATCTTAAGATCCCGTCCAGCCCT 3840

|||||

Db 8557  
CCTCAGTTTCCTCCTGTGTCCAGGAGTACACTAGATCATCTTAAGATCCCGTCCAGCCCT 8616

Qy 3841  
AAAATCATGTA CTTACTTTTTTTTTCTTTTTCTTTTTTAAATAGAGGCAAGGGTCTCTAC 3900

|||||

Db 8617  
AAAATCATGTA CTTACTTTTTTTTTCTTTTTCTTTTTTAAATAGAGGCAAGGGTCTCTAC 8676

Qy 3901  
GTTGGCCAGGCCGGTCTCAAACCTGGCCTCAAATGACTCTCCTGCCTCGGCCTCTCAA 3960

|||||

Db 8677  
GTTGGCCAGGCCGGTCTCAAACCTGGCCTCAAATGACTCTCCTGCCTCGGCCTCTCAA 8736

Qy 3961  
AGTGCTGGGATTACAGGTGTGAGCCACCGTGCCAGCTCCCTGGCCTTAAAAGTCATGTA 4020

|||||

Db 8737  
AGTGCTGGGATTACAGGTGTGAGCCACCGTGCCAGCTCCCTGGCCTTAAAAGTCATGTA 8796

Qy 4021  
ATTTAATGATCAGACCC CAGTCACAGCCATAGGATACAAAGAAGCAAAGGCAAAGAGCCC 4080

|||||

Db 8797  
ATTTAATGATCAGACCC CAGTCACAGCCATAGGATACAAAGAAGCAAAGGCAAAGAGCCC 8856

Qy 4081  
TGTGTCCTGGGCACGGTTACAGGCCAGTGTAGGGAAAGAGCTTCTGCTTGCCAGTGTGAA 4140

|||||

Db 8857  
TGTGTCCTGGGCACGGTTACAGGCCAGTGTAGGGAAAGAGCTTCTGCTTGCCAGTGTGAA 8916

Qy 4141  
GAACAGAGGAGTTTAGGAAGTGTGAGTCAGGCTCAGCTTAGTCAGGCAGAGACCAGTGGG 4200

|||||  
Db 8917  
GAACAGAGGAGTTTAGGAAGTGTGAGTCAGGCTCAGCTTAGTCAGGCAGAGACCAGTGGG 8976

Qy 4201  
CATGGGTTACCTGGGGGTAACGCGGGCCAGGTGGGCGGGCTGGCAGCCTGGGGCCCATT 4260

|||||  
Db 8977  
CATGGGTTACCTGGGGGTAACGCGGGCCAGGTGGGCGGGCTGGCAGCCTGGGGCCCATT 9036

Qy 4261  
CCTGCCAAAGCACCTCTGACCATCCTTCTCTTCACCCAGGTACTACTACAAACGGGAGAT 4320

|||||  
Db 9037  
CCTGCCAAAGCACCTCTGACCATCCTTCTCTTCACCCAGGTACTACTACAAACGGGAGAT 9096

Qy 4321  
CCTGGAACGGGTGGATGGCCGGCGACTCGTCTACAAGTTTGGCAAAAACCTCAAGCGGCTG 4380

|||||  
Db 9097  
CCTGGAACGGGTGGATGGCCGGCGACTCGTCTACAAGTTTGGCAAAAACCTCAAGCGGCTG 9156

Qy 4381  
GAAGGAGGAAGAGGTTCTCCAGAGTCGGAACCTGAGGGTTGGAACCTATACCCGGGACCAA 4440

|||||  
Db 9157  
GAAGGAGGAAGAGGTTCTCCAGAGTCGGAACCTGAGGGTTGGAACCTATACCCGGGACCAA 9216

Qy 4441  
CTCACGGACCACTCGAGGCCTGCAAACCTTCCTGGGAGGACAGGCAGGCCAGATGGCCCC 4500

|||||  
Db 9217  
CTCACGGACCACTCGAGGCCTGCAAACCTTCCTGGGAGGACAGGCAGGCCAGATGGCCCC 9276

Qy 4501  
CCCTC CCGGCTTC CCGCTC CCGCTC CCGCTC CCGCTC CCGCTC CCGCTC CCGCTC CCGCTC



TCCACTGGGGAATGCTCCCAGCTGTGCTGTGGAGAGAAGCTGATGTTTTGGTGTATTGTC 4560

Db 9277

Qy 4561

|||||

Qy 4621

|||||

Qy 4681

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	5
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	---

Oy 4741

|||||

Qy 4801

|||||

Qy 4861

[illegible]

CACGGGCAGGGGTCAGAGCACTCCCTAATTTATGTGCTATATAAATATGTCAGATGTACA 9696

Qy 4921

TAGAGATCTATTTTTTCTAAAACATTCCCCTCCCCACTCCTCTCCCACAGAGTGCTGGAC 4980

|||||

Db 9697

TAGAGATCTATTTTTTCTAAAACATTCCCCTCCCCACTCCTCTCCCACAGAGTGCTGGAC 9756

Qy 4981

TGTTCCAGGCCCTCCAGTGGGCTGATGCTGGGACCCTTAGGATGGGGCTCCCAGCTCCTT 5040

|||||

Db 9757

TGTTCCAGGCCCTCCAGTGGGCTGATGCTGGGACCCTTAGGATGGGGCTCCCAGCTCCTT 9816

Qy 5041

TCTCCTGTGAATGGAGGCAGAGACCTCCAATAAAGTGCCTTCTGGGCTTTTTCTA 5095

|||||

Db 9817

TCTCCTGTGAATGGAGGCAGAGACCTCCAATAAAGTGCCTTCTGGGCTTTTTCTA 9871

## RESULT 2

AL691482

LOCUS AL691482 112539 bp DNA linear

PRI 18-MAY-2005

DEFINITION Human DNA sequence from clone RP11-510N19 on chromosome 1 Contains

the ELF3 gene for E74-like factor 3 (ets domain transcription

factor, epithelial-specific), a novel gene, an LSM3 homolog U6

pseudogene, an small nuclear RNA associated (S. cerevisiae) (LSM3)

(ATP6V0E) ATPase H<sup>+</sup> transporting lysosomal 9kDa V0 subunit e

(CRIP1) pseudogene and a cysteine-rich protein 1 (intestinal)

pseudogene, complete sequence.

ACCESSION AL691482

VERSION AL691482.16 GI:23503795

KEYWORDS HTG; ATP6V0E; CRIP1; ELF3; LSM3.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

C C

GenCore version 6.2.1  
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2007, 04:31:56 ; Search time 854 Seconds  
(without alignments)  
10931.438 Million cell

updates/sec

Title: US-10-535-378-15  
Perfect score: 135  
Sequence: 1  
ctccgccactccggtaggat.....ccgagcaagagcgtagcctc 135

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7568541 seqs, 34560148153 residues

Total number of hits satisfying chosen parameters: 15137082

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_env:\*  
2: gb\_pat:\*  
3: gb\_ph:\*  
4: gb\_pl:\*  
5: gb\_pr:\*  
6: gb\_ro:\*  
7: gb\_sts:\*  
8: gb\_sy:\*  
9: gb\_un:\*  
10: gb\_vi:\*  
11: gb\_ov:\*  
12: gb\_htg:\*  
13: gb\_in:\*  
14: gb\_om:\*

AF110184

LOCUS AF110184 10772 bp DNA linear

PRI 22-JUL-1999

DEFINITION Homo sapiens epithelium-restricted Ets protein ESX  
gene, complete  
cds.

ACCESSION AF110184

VERSION AF110184.1 GI:5565858

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;

Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates;

Haplorrhini;

Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 4802 to 9870)

AUTHORS Chang,C.H., Scott,G.K., Kuo,W.L., Xiong,X.,  
Suzdaltseva,Y.,

Park,J.W., Sayre,P., Erny,K., Collins,C., Gray,J.W.  
and Benz,C.C.

TITLE ESX: a structurally unique Ets overexpressed early  
during human

breast tumorigenesis

JOURNAL Oncogene 14 (13), 1617-1622 (1997)

PUBMED 9129154

REFERENCE 2 (bases 1 to 10772)

AUTHORS Chang,C.H., Scott,G.K., Baldwin,M.A. and Benz,C.C.

TITLE Exon 4-encoded acidic domain in the  
epithelium-restricted Ets

factor, ESX, confers potent transactivating capacity  
and binds to

TATA-binding protein (TBP)

JOURNAL Oncogene 18 (25), 3682-3695 (1999)

PUBMED 10391676

REFERENCE 3 (bases 1 to 10772)

AUTHORS Chang,C.H., Scott,G.K. and Benz,C.C.

TITLE Direct Submission

JOURNAL Submitted (30-NOV-1998) Hematology/Oncology, U.C.S.F.,  
505

Parnassus Ave., San Francisco, CA 94143-1270, USA

FEATURES Location/Qualifiers

source

1. .10772

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

..

```

/chromosome="1"
/map="1q32"
misc_feature 34. .622
              /note="similar to THC 213038"
repeat_region 921. .1524
              /rpt_family="Alu"
              /rpt_type=dispersed
repeat_region 2978. .3293
              /rpt_family="Alu"
              /rpt_type=dispersed
CAAT_signal 4697. .4702
            /inference="non-experimental evidence, no
additional
              details recorded"
TATA_signal 4735. .4736
            /inference="non-experimental evidence, no
additional
              details recorded"
mRNA join(4777. .4888,5311. .5481,6139.
.6360,6526. .6618,
6822. .6941,7129. .7218,7364. .7480,8011.
.8206,9076. .9872)
/product="epithelium-restricted Ets protein
ESX"
5'UTR join(4777. .4888,5311. .5318)
exon 4777. .4888
      /number=1
misc_feature 4785. .4901
              /note="putative CpG island"
exon 5311. .5481
      /number=2
CDS join(5319. .5481,6139. .6360,6526.
.6618,6822. .6941,
7129. .7218,7364. .7480,8011. .8206,9076.
.9190)
      /note="epithelial-restricted with serine box;
Homo sapiens ESX cDNA ORF presented in GenBank
Accession Number U66894"
      /codon_start=1
      /product="epithelium-restricted Ets protein
ESX"
      /protein_id="AAD45237.1"
      /db_xref="GI:5565859"

```

/translation="MAATCEISNIFSNYFSAMYSSSEDSTLASVPPAATFGADDLVLTLSNPQMSLEGTEKASWLGEQPQFWSKTQVLDWISYQVEKNKYDASAI DFSRCMDMGATLCNCALEELRLVFGPLGDQLHAQLRDLTSSSSDEL SWIIELLEKDGMAFQEALDPGPFDQGS PFAQELLDDGQQASPYHPGSCGAGAPSPGSSDVSTAGTGASRSSHSSDSGGSDVDLDPTDGKLFPSDGF RDCKKGDPKHGKRKRGRPRKLSKEYWDCLEGGKSKHAPRGTHLWEFIRDILIHPELNEGLMKWENRHEGVFKFLRSEAVAQLWGQKKKNSNMTYEKLSRAMR

YYYKREILERV DGRRLVYKFGKNSSGWKEEEVLQSRN"  
repeat\_region 5773. .6059  
/rpt\_family="Alu"  
/rpt\_type=dispersed  
exon 6139. .6360  
/number=3  
exon 6526. .6618  
/number=4  
exon 6822. .6941  
/number=5  
exon 7129. .7218  
/number=6  
exon 7364. .7480  
/number=7  
misc\_feature 7401. .7525  
/note="putative CpG island"  
exon 8011. .8206  
/number=8  
repeat\_region 8655. .8775  
/rpt\_family="Alu"  
/rpt\_type=dispersed  
exon 9076. .9872  
/number=9  
3'UTR 9191. .9872  
polyA\_signal 9845. .9850  
/inference="non-experimental evidence, no  
additional details recorded"  
misc\_feature complement(9952. .10387)  
/note="similar to THC 209689"  
misc\_feature 10358. .10772  
/note="similar to THC 203540"

ORIGIN

Query Match 97.0%; Score 131; DB 5; Length 10772;

Best Local Similarity 100.0%; Pred. No. 4.3e-34;  
Matches 131; Conservative 0; Mismatches 0; Indels  
0; Gaps 0;

Qy 1  
CTCCGCCACTCCGGTAGGATTCCCCGCCTGTCATTCCCTAGCCCAGCTCTTGGGAAACTG 60

|||||  
Db 4876  
CTCCGCCACTCCGGTAGGATTCCCCGCCTGTCATTCCCTAGCCCAGCTCTTGGGAAACTG 4935

Qy 61  
CAGAGGGGTCCAGAGGATTTGCAGTTCTGAACCTGCACACTCCAGTCTAGGATCTCCGAG 120

|||||  
Db 4936  
CAGAGGGGTCCAGAGGATTTGCAGTTCTGAACCTGCACACTCCAGTCTAGGATCTCCGAG 4995

Qy 121 CAAGAGCGTAG 131  
|||||  
Db 4996 CAAGAGCGTAG 5006

RESULT 5

AL691482

LOCUS AL691482 112539 bp DNA linear

PRI 18-MAY-2005

DEFINITION Human DNA sequence from clone RP11-510N19 on  
chromosome 1 Contains

the ELF3 gene for E74-like factor 3 (ets domain  
transcription

factor, epithelial-specific), a novel gene, an LSM3  
homolog U6

small nuclear RNA associated (S. cerevisiae) (LSM3)  
pseudogene, an

ATPase H<sup>+</sup> transporting lysosomal 9kDa V0 subunit e  
(ATP6V0E)

pseudogene and a cysteine-rich protein 1 (intestinal)  
(CRIP1)

pseudogene, complete sequence.

ACCESSION AL691482

VERSION AL691482.16 GI:23503795

KEYWORDS HTG; ATP6V0E; CRIP1; ELF3; LSM3.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;

GenCore version 6.2.1  
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2007, 04:31:56 ; Search time 3358 Seconds  
(without alignments)  
10931.438 Million cell

updates/sec

Title: US-10-535-378-11  
Perfect score: 531  
Sequence: 1  
ccgggctgagcggcttcctg.....acctgtgggagttcatccgg 531

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7568541 seqs, 34560148153 residues

Total number of hits satisfying chosen parameters: 15137082

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_env:\*  
2: gb\_pat:\*  
3: gb\_ph:\*  
4: gb\_pl:\*  
5: gb\_pr:\*  
6: gb\_ro:\*  
7: gb\_sts:\*  
8: gb\_sy:\*  
9: gb\_un:\*  
10: gb\_vi:\*  
11: gb\_ov:\*  
12: gb\_htg:\*  
13: gb\_in:\*  
14: gb\_om:\*



Db 775  
AGGCAGCTGGTGGGTCTTAGGTGGGCTGAGGAGAAAAGCAGTCACTGCAGTACCCGCACA 834

Qy 361  
GAGGGCACTGCGGGGTCTCTGGAGAGGCTTGCTGCATGCTGTGGCCAAGTCAAGCAGTGC 420  
|||||

Db 835  
GAGGGCACTGCGGGGTCTCTGGAGAGGCTTGCTGCATGCTGTGGCCAAGTCAGCAGTGCA 894

Qy 421  
ACTGGGGCGGCAGGGCTGGCTGGCCTTGGGTGAGAGGGGGCACCTGGATGGCAAACGGAT 480  
||| |||||  
||||| |||

Db 895  
CTGGGGCGGGCAGGGCTGGCTGGCCTTGGGTGAGAGGGGACACCTGGATGGCAAACCTGAT 954

Qy 481 GGAGGCTGG-CTTGCAGCGCCCAGAGGCACCCACCTGTGGGAGTTCATCCGG  
531  
||||| |||||

Db 955 GGAGGCTGGCCTTGCAGCGCCCAGAGGCACCCACCTGTGGGAGTTCATCCGG  
1006

# RESULT 2

AF110184

LOCUS AF110184 10772 bp DNA linear

PRI 22-JUL-1999

DEFINITION Homo sapiens epithelium-restricted Ets protein ESX gene, complete

cds.

ACCESSION AF110184

VERSION AF110184.1 GI:5565858

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;

Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates;

Haplorrhini;

Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 4802 to 9870)

AUTHORS Chang,C.H., Scott,G.K., Kuo,W.L., Xiong,X.,  
Suzdaltseva,Y.,

Park,J.W., Sayre,P., Erny,K., Collins,C., Gray,J.W.  
and Benz,C.C.

TITLE ESX: a structurally unique Ets overexpressed early  
 during human breast tumorigenesis  
 JOURNAL Oncogene 14 (13), 1617-1622 (1997)  
 PUBMED 9129154  
 REFERENCE 2 (bases 1 to 10772)  
 AUTHORS Chang,C.H., Scott,G.K., Baldwin,M.A. and Benz,C.C.  
 TITLE Exon 4-encoded acidic domain in the  
 epithelium-restricted Ets  
 factor, ESX, confers potent transactivating capacity  
 and binds to TATA-binding protein (TBP)  
 JOURNAL Oncogene 18 (25), 3682-3695 (1999)  
 PUBMED 10391676  
 REFERENCE 3 (bases 1 to 10772)  
 AUTHORS Chang,C.H., Scott,G.K. and Benz,C.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-NOV-1998) Hematology/Oncology, U.C.S.F.,  
 505 Parnassus Ave., San Francisco, CA 94143-1270, USA  
 FEATURES Location/Qualifiers  
 source 1. .10772  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="1q32"  
 misc\_feature 34. .622  
 /note="similar to THC 213038"  
 repeat\_region 921. .1524  
 /rpt\_family="Alu"  
 /rpt\_type=dispersed  
 repeat\_region 2978. .3293  
 /rpt\_family="Alu"  
 /rpt\_type=dispersed  
 CAAT\_signal 4697. .4702  
 /inference="non-experimental evidence, no  
 additional details recorded"  
 TATA\_signal 4735. .4736  
 /inference="non-experimental evidence, no  
 additional details recorded"  
 mRNA join(4777. .4888,5311. .5481,6139.  
 .6360,6526. .6618,  
 6822. .6941,7129. .7218,7364. .7480,8011.

```

.8206,9076. .9872)
ESX" /product="epithelium-restricted Ets protein
5'UTR join(4777. .4888,5311. .5318)
exon 4777. .4888
/number=1
misc_feature 4785. .4901
/note="putative CpG island"
exon 5311. .5481
/number=2
CDS join(5319. .5481,6139. .6360,6526.
.6618,6822. .6941,
7129. .7218,7364. .7480,8011. .8206,9076.
.9190)
/note="epithelial-restricted with serine box;
Homo sapiens ESX cDNA ORF presented in GenBank
Accession Number U66894"
/codon_start=1
/product="epithelium-restricted Ets protein
ESX"
/protein_id="AAD45237.1"
/db_xref="GI:5565859"

/translation="MAATCEISNIFSNYFSAMYSSDSTLASVPPAATFGADDLVLT
SNPQMSLEGTEKASWLGEQPQFWSKTQVLDWISYQVEKNKYDASAI DFSRCMDGATL
CNCALEELRLVFGPLGDQLHAQLRDLTSSSSDELSWIIELLEKDGMAFQEALDPGPF
QGSPFAQELLDDGQQASPYHPGSCGAGAPSPGSSDVSTAGTGASRSSHSSDSGGSDVD
LDPTDGKLFPSDGF RDCKKGD PKHGKRKRGRPRKLSKEYWDCLEGGKSKHAPRGTHLW
EFIRDILIHPELNEGLMKWENRHEGVFKFLRSEAVAQLWGQKKKNSNMTYEKLSRAMR
YYYKREILERV DGRRLVYKFGKNSSGWKEEEVLQSRN"
repeat_region 5773. .6059
/rpt_family="Alu"
/rpt_type=dispersed
exon 6139. .6360
/number=3
exon 6526. .6618
/number=4
exon 6822. .6941
/number=5

```

```

exon          7129. .7218
               /number=6
exon          7364. .7480
               /number=7
misc_feature  7401. .7525
               /note="putative CpG island"
exon          8011. .8206
               /number=8
repeat_region 8655. .8775
               /rpt_family="Alu"
               /rpt_type=dispersed
exon          9076. .9872
               /number=9
3'UTR        9191. .9872
polyA_signal  9845. .9850
               /inference="non-experimental evidence, no
additional
               details recorded"
misc_feature  complement(9952. .10387)
               /note="similar to THC 209689"
misc_feature  10358. .10772
               /note="similar to THC 203540"

```

# ORIGIN

```

Query Match          91.9%;  Score 488;  DB 5;  Length 10772;
Best Local Similarity 96.1%;  Pred. No. 9.5e-139;
Matches 511;  Conservative 0;  Mismatches 20;  Indels
1;  Gaps 1;

```

```

Qy          1
CCGGGCTGAGCGGCTTCCTGGGGCACTGCGGGTTGTTGCAGGTATCCCCTCTCCCGTTTC 60
      |||
| | | | |
Db          7514
CCGGGCTGAGCGGCTTCCTGGGGCACTGCGGGTTGTTGCAGGTATCCCTTCTCCCGTTTT 7573

```

```

Qy          61
CTCTGGCCTCCGCATGGCCTTTGGTAAGGCTGTGCACAAGCTGGGGGCTCTATGGTATCG 120
      |||
| | | | |
Db          7574
CTCTGGCCTCCGCATGGCCTTTGGTAAGGCTGTGCACAAGCTGGGGGCTCTATGGTATCG 7633

```

```

Qy          121
GTCACCACCTAATTGCAGAGCCAGGCTTGGTGGTCCTGGAGAGGAGGAGGAAATAAGGCT 180
      |||
.....

```

|||||

Db 7634

GTCACCACCTAATTGCAGAGCCTGGCTTGGTGGTCCTGGAGAGGAGGAGGAAATAAGGCT 7693

Qy 181

CCCAGTGGGAGGCTCATGGTACCAGAGTCCTGTCCACTGACTCCAGTGTCTGTCCACTG 240

|||||

Db 7694

CCCAGTGGGAGGCTCATGGTACCAGAGTCCTGTCCACTGACTCCAGTGTCTGTCCACTG 7753

Qy 241

ACTCCAGTTCTCTCTGCACTTGGCCACTGTCCTGCCCTCTGGGTCACCCTCAATGTGAGG 300

|||||

|||||

Db 7754

ACTCCAGTTCTCTCTGCACTTGGCCACTGTCCTGCCCTCTGGGACACCCTCAATGTGAGG 7813

Qy 301

AGGCGGCTGGTGGGTCTTAGGTGGGCTGAGGAGAAAAGCAGTCACTGCAGTACCCGCACA 360

||||

|||||

Db 7814

AGGCGAGCTGGTGGGTCTTAGGTGGGCTGAGGAGAAAAGCAGTCACTGCAGTACCCGCACA 7873

Qy 361

GAGGGCACTGCGGGGTCTCTGGAGAGGCTTGCTGCATGCTGTGGCCAAGTCAAGCAGTGC 420

|||||

Db 7874

GAGGGCACTGCGGGGTCTCTGGAGAGGCTTGCTGCATGCTGTGGCCAAGTCAAGCAGTGC 7933

Qy 421

ACTGGGGCGGCAGGGCTGGCTGGCCTTGGGTGAGAGGGGGCACCTGGATGGCAAACGGAT 480

||| |||||

|||||

Db 7934

CTGGGGCGGGCAGGGCTGGCTGGCCTTGGGTGAGAGGGGACACCTGGATGGCAAACCTGAT 7993

Qy 481 GGAGGCTGG-CTTGCAGCGCCCAGAGGCACCCACCTGTGGGAGTTCATCCGG  
531

|||||

Db 7994 GGAGGCTGGCCTTGCAGCGCCCAGAGGCACCCACCTGTGGGAGTTCATCCGG  
8045